

ABSTRACT

A method of extracting significant phrases from one or more documents stored in a computer-readable medium. A sequence of words is read from the one or more documents and a score is determined for each word in the sequence based on the length of the word. The score for each word in the sequence is compared against a threshold score. The sequence of words is indicated to be a significant phrase if the number of words in the sequences that have a score greater than the threshold score equals or exceeds a predetermined number. A sentence containing the sequence of words is retrieved from the document, if the sequence of words is a significant phrase. An abstract of the document is searched to determine if the sentence has been previously included in the abstract. If not, the sentence is added to the abstract.

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100